Indonesian Journal of English Language Teaching
Volume 4/Number 2 • October 2008

LEXICAL PROFILES IN EAP SPEAKING TASK PERFORMANCE

Noriko Iwashita
The University of Queensland

Abstract

The present study investigates lexical competence in performance on speaking tests. It examines the extent to which learners preparing for tertiary study in English-speaking countries are able to demonstrate their ability to use a wide range of vocabulary in carrying out academic speaking tasks. Ninety-six task performances over four different tasks and two task types were drawn from three different levels. The performances were transcribed and analyzed using the WordSmith program (Scott, 2004). The results showed that test-takers' vocabulary varied according to task and task type. The results of the study have implications for task design in academic speaking tests and teaching/learning vocabulary in EAP courses.

Keywords: Lexical competence; Academic speaking test; Task types; EAP courses

INTRODUCTION

The examination of English proficiency has become an important means of providing tertiary institutions with precise information on students' competence in handling the academic English needed for understanding lectures, participating in class discussions and writing essays. This is because the number of non-native speakers of English studying at universities in English-speaking countries has increase tremendously. Thus there is a need for students to be able to use a wide range of vocabulary in an academic setting in order to satisfy the requirements of academic study. The proposed study investigates how ESL learners preparing for tertiary study in an English-speaking country demonstrate lexical knowledge in their performance on English for Academic Purposes speaking test. In particular,

Direct all correspondence to: School of Languages and Cultural Studies, The University of Queensland, Brisbane, Queensland Australia, 4072. n.iwashita@uq.edu.au <mailto:n.iwashita@uq.edu.au> Phone. 61-73365 2381 Fax. 61-7-3365 6799.

the study aims to examine whether different task types would influence learners' choice of words in performing their task(s). Drawing on the quantitative linguistic analysis of the MICASE (Michigan Corpus of Academic Spoken English) (Simpson, Briggs, Ovens, & Swales, 2003), the present study involved a comprehensive linguistic description of the range of spoken registers used by ESL learners in their academic speaking test performance on two different task types.

BACKGROUND

Investigation into the use of vocabulary in the performance of academic tasks has been studied mainly in the contexts of writing and reading. For example, Enober (1995) examined the lexical component as one factor in holistic scoring. Sixty-six essays by non-native speakers of English from various language backgrounds were holistically scored compared with four lexical richness measures (lexical variation, error-free variation, percentage of lexical error, and lexical density). The results showed significant high correlations for lexical variation. Santos (1988) investigated the reactions of academics to essays by 96 students who were native speakers of Korean or Chinese and found that vocabulary errors were regarded as the most serious. Leki and Carson (1994) conducted a survey asking non-native English-speaking students about what they would like to learn in EAP courses, and found that vocabulary was identified as the first priority. Although a considerable amount of research is devoted to the role of vocabulary knowledge in academic writing, little is known about how important it is for learners to possess a wide range of vocabulary in academic speaking.

Academic speech is defined as speech that occurs in academic settings and includes both rehearsed and spontaneous speech (e.g., Lindemann & Mauranen, 2001). Most studies have been devoted to investigate the characteristics of academic discourse such as rhetorical organization of classroom discourse and lectures, or to examining registers frequently observed in academic speech in corpus-based study (e.g., Lindemann & Mauranen, 2001). Few, however, have investigated lexical competence in test performance in relation to the use of a variety of vocabulary in carrying out academic tasks such as discussions and presentations.

Proficiency tests such as the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS) provide university administrators with information about whether test-takers are able to cope with tertiary study in English-speaking countries. In the

speaking component of such tests, tasks are designed to simulate situations that test-takers are likely to use in an academic context. Brown, McNamara and Iwashita (2005) examined comments on academic speaking performance by expert EAP teachers in the context of scale development, and found that EAP teachers made general assessments of test-takers' vocabulary skills and frequently commented on the adequacy of their vocabulary for a particular task. Furthermore, analysis of test-taker discourse has shown that lexical knowledge is one of the most important features distinguishing proficiency levels of examinees (e.g., Iwashita, Brown, McNamara & O'Hagan, 2008), but most research has examined general vocabulary (e.g., Douglas, 1994) rather than the specific lexis of academic spoken English. Iwashita et al., (2008) investigated the use of academic vocabulary in speaking test performances using the Academic Word List (Coxhead, 1998), and found little difference across task types. The Academic Word List (AWL) was compiled from a corpus of 3.5 million running words of written academic text, i.e. not from spoken corpora. Some recent studies identified typical features of academic speech (e.g., Camiciottoli, 2004) which are not observed in academic written English. For this reason, it is important to investigate how lexical profiles in academic spoken English may differ according to proficiency levels and task types based on the academic spoken corpus.

Speaking tasks in EAP tests increasingly seek to replicate the roles of and demands on students in the academic context. Integrated tasks, in which test-takers are required to process and transform cognitively complex stimuli (written texts, lectures etc.) and integrate the information into their speaking performance are widely used in academic speaking tests. Integrated tasks are more complex and demanding than traditional independent tasks, where test-takers draw on their own knowledge or ideas in response to a question or prompt, and where the absence of input means that the tasks are often restricted to fairly bland topics drawing on the test-takers' general knowledge. According to Skehan (1998), performance on integrated tasks is generally less accurate and fluent than on independent tasks. However, producing speech using the information presented in the prompt may enhance the quality of lexical aspects of performance, but so far this aspect has not been researched.

RESEARCH QUESTION

The present study addresses the following research question:

To what extent do learners' vocabulary differ according to task and task type in academic speaking?

METHODOLOGY

Data

Do include the fact that data is drawn fom a corpus from MICASE. Explain wjjat is MICASE and provide a rationale for using MICASE for your data.

Comment [NI1]: Actually the data is not drawn from MICASE. For analysis of the data, we used MICASE as a reference. It is explained on page 6

The data used for the study were initially collected in the USA as part of the piloting of materials in the development of the next generation of TOEFL. For the purposes of this project, ten samples of each task at each of the upper three levels (levels 3–5) were randomly selected from a larger pool of pilot test data: a total of 24 performances per task and 96 in total. The ESL learners who took the trial test varied in age, L1, length of residence in an English-speaking country and prior time spent studying English, but all were studying English to prepare for tertiary study in the USA at the time of data collection.

Tasks

The four test tasks used in the present study were of two types, independent and integrated. This was based on whether performance involved prior comprehension of extended stimulus materials. In the independent tasks, participants were asked to express their opinion on a certain topic, which was presented with no accompanying material to read or hear. In the integrated tasks, participants first listened to or read information presented in the prompt, and then were asked to explain, describe or recount the information. The amount of preparation and speaking time varied for each task, but longer preparation and speaking times were given for the integrated tasks than for the independent ones (Table 1).

TABLE 1
The academic speaking tasks

Task	Type	Targeted functions and discourse features	Preparation time (secs)	Speaking time (secs)
1	Independent	Opinion; Impersonal focus; Factual/conceptual information	30	60
2	Independent	Value/significance; Impersonal focus; Factual/conceptual information	30	60
3	Integrated;	Explain/describe/recount;	60	90

	Monologic lecture	Example/event; Cause/effect		
4	Integrated; Reading	Explain/describe/recount; Process/procedure; Purpose/results	90	90

Data analysis

Data were analyzed in two stages. Stage 1 involved general vocabulary use (i.e., vocabulary use which context is not specified) and stage 2 involved the use of vocabulary specific to context and academic speech. Prior to analysis, the transcribed speech was pruned to exclude features of repair and imported into VocabProfile (Cobb, 2002). Word type and token were counted. In order to enable comparisons across tasks with different times allowed for completion, instances of word-token and word-type were counted per 60 seconds of speech. It should be noted that since a limited speaking time is allowed for each task, test-takers who can speak fast produce more word-tokens than test-takers who speak slowly. For that reason, the number of word-tokens could be affected by the speed of test-taker speech.

To examine learners' ability to use vocabulary specific to the task context and to academic speech, the data were further analyzed using WordSmith Tools (Scott, 2004) and MICASE . WordSmith Tools is an integrated suite of programs for looking at how words behave in text. The KeyWords program in the tools was used to identify the keywords in the text. Keywords are those whose frequency is unusually high in comparison with some norm but are not the most frequent words (Scott, 2004). The keywords were calculated by comparing the frequency of each word in the word list of the transcribed performances with the frequency of the same word in the reference word list. In the present study, MICASE was used as the reference list. The MICASE corpus is a spoken language corpus of approximately 1.8 million words (200 hours) focusing on academic speech collected in the University of Michigan. MICASE was chosen as a reference list as it is a corpus of academic speech, which is compatible with the type of speech collected for the present study. It was expected that if learners used words typically used in academic settings, the KeyWords program would identify both context-dependent and context-independent words (note: context dependent words such as proper nouns were excluded from the analysis). The results of the analysis are reported with the number and type of keyword tokens and also the percentages of keyword token and type in the total number of word tokens and types. The effects of task and task type (i.e., independent and integrated) on these measures were examined using inferential statistics that is MANOVA and T-tests.

RESULTS AND DISCUSSION

Tables 2 and 3 present descriptive statistics of general vocabulary and keywords according to four different tasks, and include both raw and frequency data. It was assumed that learners would produce more words in Tasks 3 and 4, as the required speaking time was longer than in Tasks 1 and 2. However, as shown in the frequency data, the independent tasks produced more words per 60 seconds than did the integrated tasks (Table 2). However, instances of keywords were observed far more frequently in integrated task performances (Tasks 3 and 4) than in independent task performances (Tasks 1 and 2) for all four measures.

TABLE 2
Descriptive statistics of vocabulary use (per task)

	Token		Type		Token/60secs		Type/60 secs	
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	105.63	27.76	62.71	12.59	116.09	28.21	69.58	15.67
2	97.08	25.69	60.00	12.12	104.59	29.40	64.81	15.61
3	156.00	36.45	79.83	18.17	90.18	19.42	46.35	10.81
4	117.62	29.81	64.79	14.67	86.67	23.03	47.82	11.92

TABLE 3
Descriptive statistics of key words and percentages (per task)

	Keyword token		Keyword type		Token (%)		Type (%)	
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	9.42	6.636	2.16	1.385	0.08	0.04	0.03	0.02
2	8.07	6.364	2.4	1.454	0.08	0.05	0.04	0.03
3	13.17	10.499	3.79	1.744	0.12	0.05	0.07	0.03
4	18.46	7.616	5.13	2.173	0.11	0.08	0.06	0.03

The results of multivariate analysis show that the effect of task on vocabulary measures was found to be significant with small effect size (F [12, 197] = 4.58 p = 0.001 *eta* = 0.22). Task-type comparisons are summarized in Table 4. As for task-type comparisons, all measures except word token were found to be significantly different between independent and integrated tasks (General vocabulary) .

Comment [J2]: This is unclear . Can you explain further.

Comment [NI3R2]: General vocabulary is explained on page 5.

TABLE 4
Comparison of task types

	Task type	*M*	*SD*	*t*	*df*	Sig.
General vocabulary						
Token	1	104.03	26.08	-7.2	70	0.001
	2	156.00	36.45			
Type	1	61.85	12.28	-5.10	70	0.001
	2	79.83	18.17			
Token/60secs	1	111.88	28.49	3.06	70	0.003
	2	90.18	19.42			
Type/60secs	1	66.93	15.92	5.85	70	0.001
	2	46.35	10.81			
Keyword analysis						
KWToken	1	8.82	6.46	-1.95	56	0.056
	2	13.17	10.50			
KWtype	1	2.26	1.40	-3.70	56	0.001
	2	3.79	1.74			
KWToken (%)	1	0.08	0.05	-3.02	56	0.004
	2	0.12	0.05			
KWtype (%)	1	0.04	0.02	-4.22	56	0.001
	2	0.07	0.03			

Notes: Task type 1 – independent 2 – integrated task; KWToken – Keyword token, KWtype – Keyword type

The results presented in table 4 above show that the number of keywords varied according to task and task type. Similar results were obtained by analyzing the data in the present study using the Academic Word List (Brown et al., 2005). Brown et al. investigated the percentage of words in each of the four categories: the most frequent 1,000 English words, the second most frequent 1,000 English words, words in the Academic Word List, and any remaining words. This was done using the *VocabProfile* program (Cobb, 2002), which is based on the Vocabulary Profile (Laufer & Nation, 1995) and the Academic Word List (Coxhead, 1998). The results are summarized in Tables 5 and 6 below. The number of academic word tokens was larger in the integrated task performances (Tasks 3 and 4), but this was explained by the longer speaking time required for those tasks. The percentage of academic vocabulary was not large in the integrated tasks, which conflicts with the findings of the present study.

TABLE 5
Descriptive statistics of academic words (per task)

	Token		%	
Task	*M*	*SD*	*M*	*SD*
1	3.92	1.79	3.87	2.19
2	3.58	3.27	3.51	3.01
3	5.50	2.74	3.55	1.51
4	4.29	3.13	3.67	2.71

TABLE 6
Comparison of the use of academic words between task types

	Task type	Mean	Std. Deviation	t	df	Sig. (2-tailed)
Token	1	3.44	1.85	-2.638	70	0.01
	2	5.50	2.74			
Token %	1	3.30	1.82	0.251	70	0.80
	2	3.55	1.51			

Task type: 1 – independent; 2 – integrated

The comparison of the analysis of different types of corpora (KeyWords analysis and Academic Word List analysis) could be explained by the words identified as keywords against MICASE using the KeyWords program and the academic words identified in the AWL (Coxhead, 1998). The following transcripts of test-taker performances show the words identified in both KeyWords and the Academic Word List (note: underlined words in bold were identified by both AWL and KeyWords; italicized words were identified by the KeyWords Program; words in bold were identified by the AWL).

Example 1 (Independent task Level 3)
I think that music and art could be encourage because this course can active the *children* **creativity** and for nature children are **creative** and when are children let the people if people develop their **creativity** when they are children they will be able to perform better when well be **adults** so in this sense they can be more productive they can **create** more they can be more helpful to the society and to the companies

Example 2 (Integrated task Level 5)
The San Juaquin Valley presented as a place where *land* **subsidence occurred.** The San Juaquin Valley **located** in California was using *groundwater* from the late eighteen eighties. Now there was heavy *pumping* of water for both irrigation and other purposes in this valley. By the twenties

and thirties land **subsidence** had already **occurred** and by the early nineteen seventies because of the unabated use of groundwater *groundwater* levels had *sunk* by hundred and twenty metres while the land had dropped by a level of eight metres. Now this might seem like a large amount but it **occurred** over a long **period** of time. So in order to mitigate this problem in the nineteen seventies. San Juaquin Valley reduced *pumping* of water and increased the use of surface water however the problem of land **subsidence** reappeared in the nineteen nineties because of the drought in California. And this made people start using groundwater again. And it was even a huger problem now because groundwater levels sunk by much greater than the seventies and the land level sunk greatly too.

As shown in the examples above, the academic words identified in the AWL analysis and the words identified in the KeyWords program using the MICASE as a reference list are somewhat different. Words such as 'occurred' 'subsidence' and 'creativity' were identified in both analyses. That is, they are listed in the AWL and are also frequently observed according to the KeyWords analysis. Words such as 'groundwater' and 'pumping' were identified in the KeyWords analysis, but are not listed in the AWL. Many words listed in the AWL and registers specific to academic speech were not identified in the KeyWords analysis. This does not mean that learners did not use academic words (they did, as is shown in the results of the AWL analysis) or academic registers specific to speaking, but that the words might not occur frequently enough to be captured by the KeyWords analysis. In order to examine whether learners use registers specific to academic speech, we need first to identify the types of registers used in academic speech using MICASE, as in other studies.

Comment [J4]: Please explain in what way they are different/

Comment [NI5R4]: This is explained in the subsequent sentences.

CONCLUSION

The present study shows a lexical profile of learners in their academic speaking-test performances. Also, the comparison of the results from the KeyWords and AWL analyses provides useful information on the use of integrated tasks in academic speaking tests. As was discussed earlier, it was assumed that cognitively demanding integrated tasks would produce more sophisticated speech in terms of grammatical and lexical complexity. However, as shown in the results of the AWL analysis, academic words from the academic written corpus were used more frequently in independent tasks than in integrated tasks, but, according to the KeyWords analysis, context-specific words were produced more frequently in the integrated tasks than in the independent tasks. Because an input text was given as a prompt, it was assumed that learners used content words from the text far

more frequently in the integrated tasks in summarizing and recounting what they had heard. However, integrated tasks do not necessarily provide opportunities for learners to demonstrate the knowledge of academic vocabulary necessary for academic speaking. To examine academic vocabulary use it would be better to employ independent rather than integrated tasks, but it is still not clear how learners use (or whether they actually use) academic registers specific to academic speaking. Further investigation will be required to understand the fine details of lexical profiles in academic speaking-task performances.

ACKNOWLEDGEMENTS

The research project reported in this paper was awarded a 2004 Spaan Fellowship by the English Language Institute, The University of Michigan. The data examined in the study were drawn from a large project funded by Educational Testing Service (ETS) for the work of the TOEFL Speaking Team. I would like to express my gratitude to Annie Brown and Tim McNamara for allowing me to use the data for this project and also their contribution to the earlier study. I am also grateful to Gavin Melles for his suggestions on data analysis.

THE AUTHOR

Noriko Iwashita is assistant professor in applied linguistics at the University of Queensland, Brisbane Australia. Her research interests include analysis of interlanguage, interactionist accounts of second language learning, and task-based assessment. Her work has appeared *in Studies in Second Language Acquisition, Applied Linguistics, Language Learning, Language Testing, Acquisition of Japanese as a Second Language and System.*

REFERENCES

Brown, A., N. Iwashita, & T. McNamara (2005). *An examination of rater orientations and test-taker performance on English for Academic Purposes speaking tasks.* TOEFL Monograph Series. Princeton, NJ: Educational Testing Service.

Camiciottoli, B. C. (2004). Interactive discourse structuring in L2 guest lectures: Some insights from a comparative corpus-based study. *Journal of English for Academic Purposes 3*, 39–54.

Cobb, T. (2002). *The Web vocabulary profiler*. (Available from http://www.er.uqam.ca/nobel/r21270/textools/web_vp.html).

Coxhead, A. J. (1998). *An academic word list*. Wellington, New Zealand: Victoria University of Wellington.

Douglas, D. (1994). *Quantity and quality in speaking test performance. Language Testing 11* (2), 125–144.

Enober, C. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing 4* (2), 139–155.

Iwashita, N., Brown, A., McNamara, T. & O'Hagan, S. (2008). What features of language distinguish levels of learner proficiency? in-depth analysis of task performance in the context of the speaking scale development. *Applied Linguistics 29* (1) 24-49.

Leki, I. & J. Carson (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly, 28* (1), 81–101.

Laufer, B. & P. Nation (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics 16* (3): 307–322.

Lindemann, S. & A. Mauranen (2001). *It's just real messy:* The occurrence and function of just in a corpus of academic speech. *English for Specific Purposes 20,* 459–475.

Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly 32* (1), 69–91.

Scott, M. (2004). WordSmith Tools (version 4) [computer software]. Oxford: Oxford University Press. (Available from http://www.lexically.net/wordsmith/).

Simpson, R. C., S. L. Briggs, J. Ovens, & J. M. Swales (2003). *The Michigan corpus of academic spoken English.* Ann Arbor: The University of Michigan.

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford: Oxford University Press.